



# Машинное обучение для решения исследовательских и инженерных задач в науках о Земле

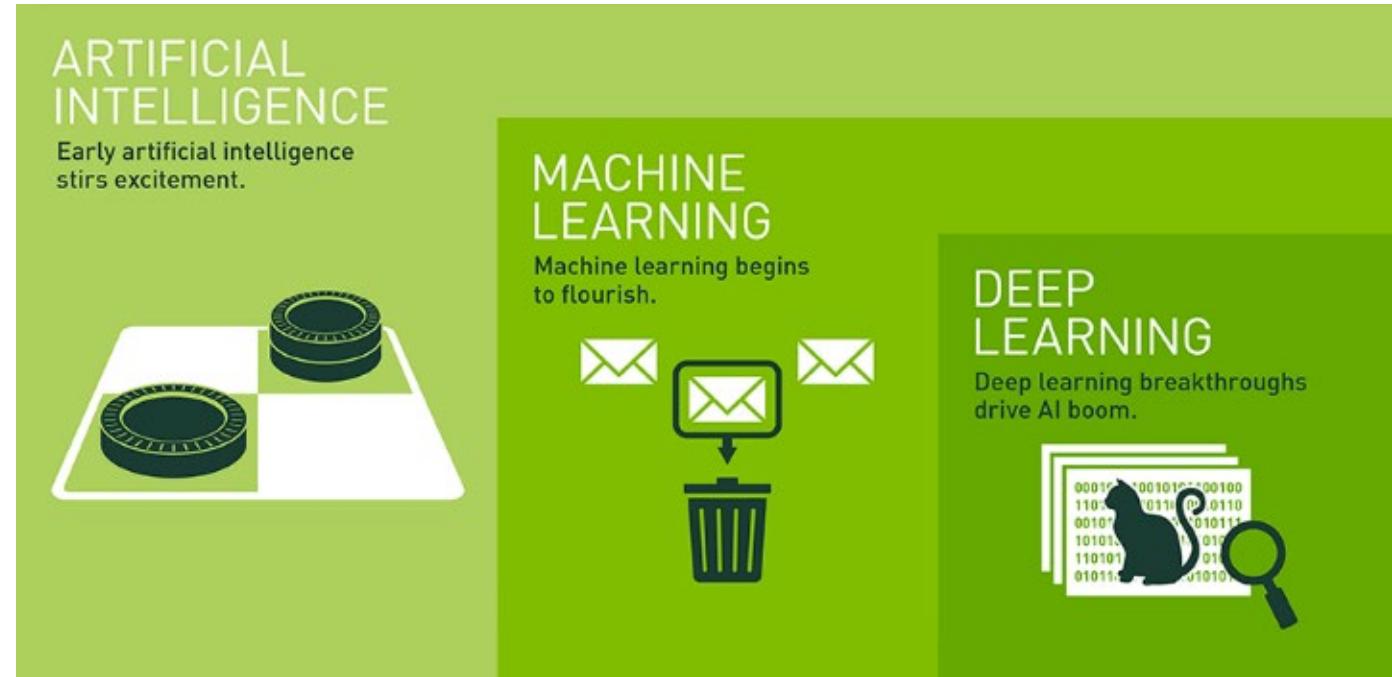
Михаил Криницкий

к.т.н., н.с.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и  
мониторинга климатических изменений (ЛВОАМКИ)

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ, ГЛУБОКОЕ ОБУЧЕНИЕ

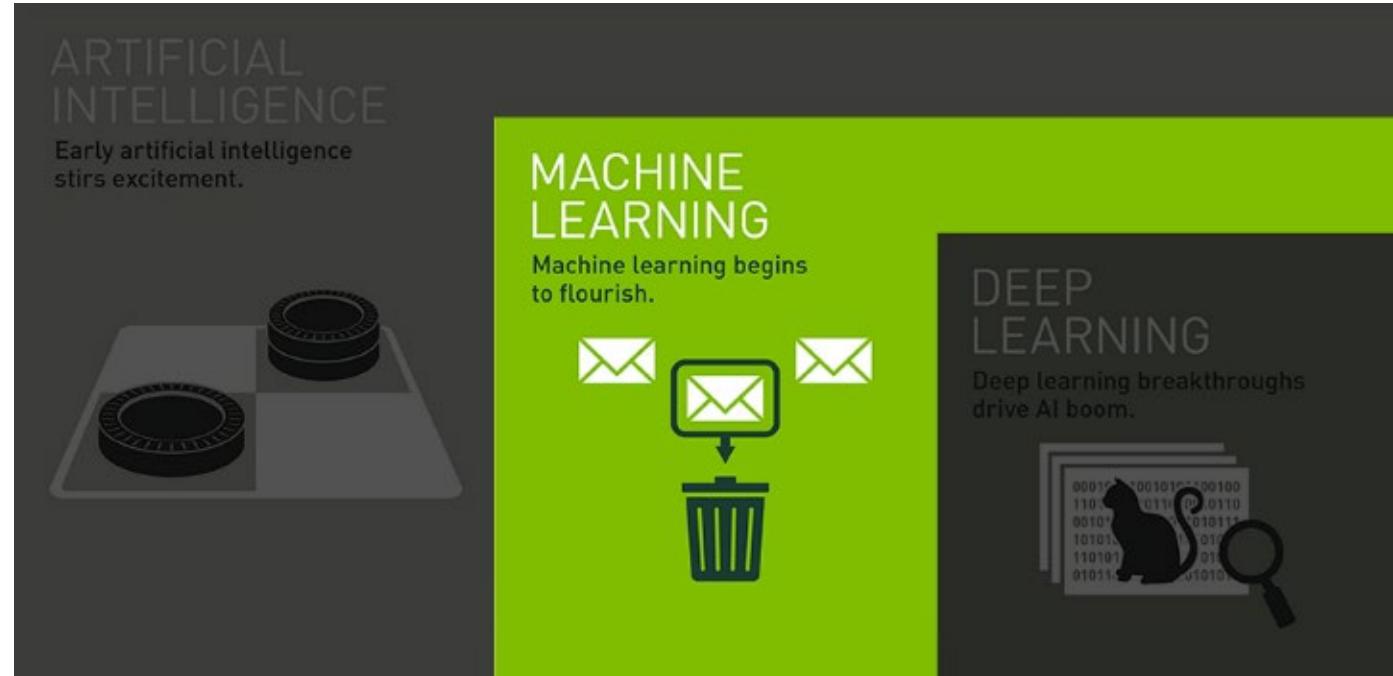


**Искусственный интеллект (ИИ)** - это наука и инженерная технология создания интеллектуальных машин, и в особенности интеллектуальных компьютерных программ. ИИ связан со сходной задачей использования компьютеров для понимания человеческого интеллекта, но не обязательно ограничивается биологически правдоподобными методами.

Дж. МакКарти, 1956г.

**ИИ** - научное направление, в рамках которого **ставятся и решаются задачи** аппаратного или программного **моделирования** тех **видов человеческой деятельности**, которые **традиционно считаются интеллектуальными**

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ, ГЛУБОКОЕ ОБУЧЕНИЕ

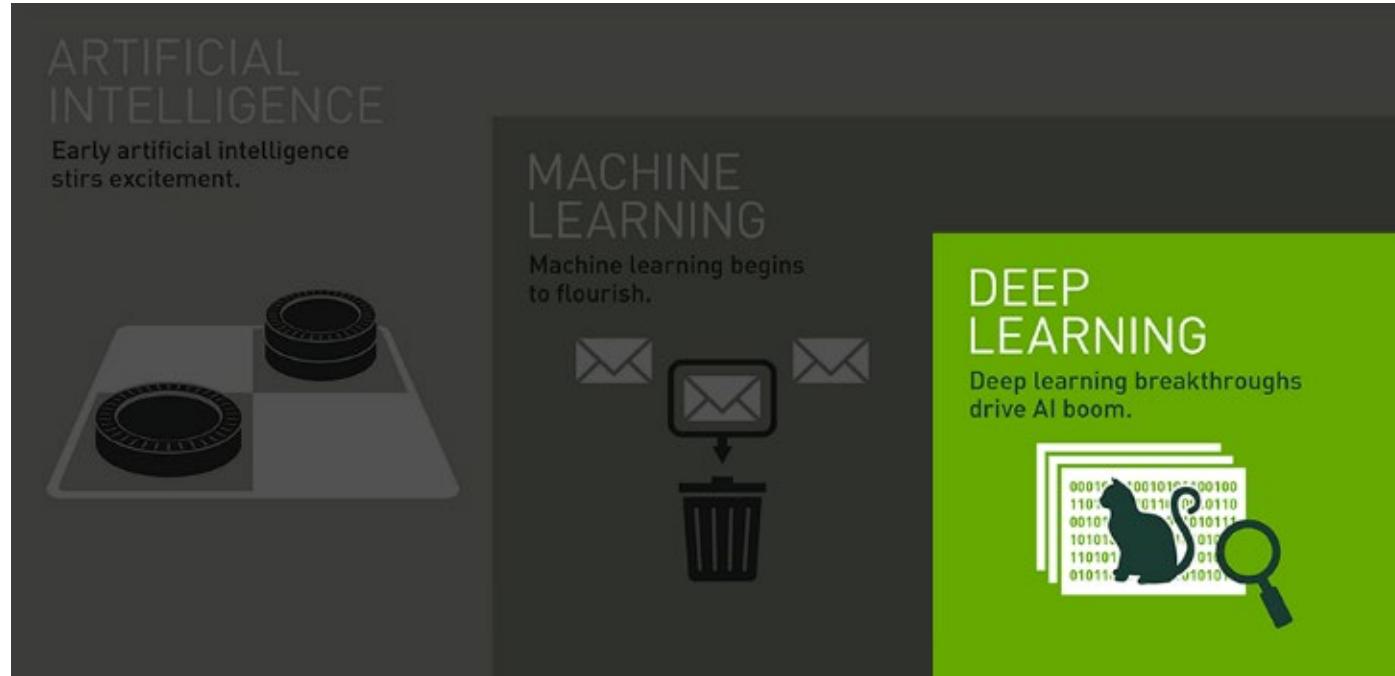


## Машинное обучение (МО)

«область компьютерных наук, придающих способность компьютерам обучаться без необходимости явно их программировать»<sup>(1)</sup>

<sup>(1)</sup>A. L. Samuel, Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, vol. 3, no. 3, p.p. 210-229, July 1959. doi: 10.1147/rd.33.0210

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ, ГЛУБОКОЕ ОБУЧЕНИЕ



## Глубинное обучение (Deep Learning, DL)

Совокупность **методов машинного обучения**, основанных на обучении представлениям данных, а не специализированным алгоритмам, предназначенным для решения конкретных задач.



# Классификация задач и методов машиинного обучения

Михаил Криницкий

к.т.н., н.с.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и  
мониторинга климатических изменений (ЛВОАМКИ)

# ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ

Когда (человеку) непонятно, что происходит  
**все равно строим модель**

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?

# ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ

Когда (человеку) непонятно, что происходит  
**все равно строим модель**

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы)  $\{y_i\}$
- 
- 
-

# ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ

Когда (человеку) непонятно, что происходит  
**все равно строим модель**

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы)  $\{y_i\}$
- Данные измерений  $\{x_i\}$  (желательно  
релевантных задаче)

# ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ

Когда (человеку) непонятно, что происходит  
**все равно строим модель**

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы)  $\{y_i\}$
- Данные измерений  $\{x_i\}$
- $\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$

# ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ

Когда (человеку) непонятно, что происходит  
**все равно строим модель**

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы)  $\{y_i\}$
- Данные измерений  $\{x_i\}$
- $\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$
- Обобщающая. Достоверная (в каком смысле?)

Применимая.

# ПРИМЕНЕНИЕ СОВРЕМЕННЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ НАУК О ЗЕМЛЕ

## классический подход

- Трактуемость (интерпретируемость)
- В сложных случаях – долгая реализация
- Меняющиеся условия и неучтенные воздействия снижают точность
- Подход основан на понимании описываемого процесса. Модель процесса от начала до конца формулируется человеком.  
поэтому нередко модель **недостаточно сложна** для описания сложного процесса с **желаемой точностью**.

## методы машинного обучения

- Трактуемость (интерпретируемость) – **редко**  
ММО не всегда подходят для установления причинно-следственных зависимостей
- Скорость реализации
- Модели можно «дообучать» в меняющихся условиях
- Подход основан на данных. Человек может участвовать (помогать) на этапе выбора типа модели общего вида.  
модели МО обладают широким спектром сложности, что **позволяет аппроксимировать** очень сложные процессы с **хорошей точностью**.

# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

!!! ЦЕЛЬ: строим модель для решения задачи !!!

# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: строим модель для решения задачи

Балансируем:

интерпретируемость

точность («качество»)

способность к обобщению

энерго-, вычислительные затраты

# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)

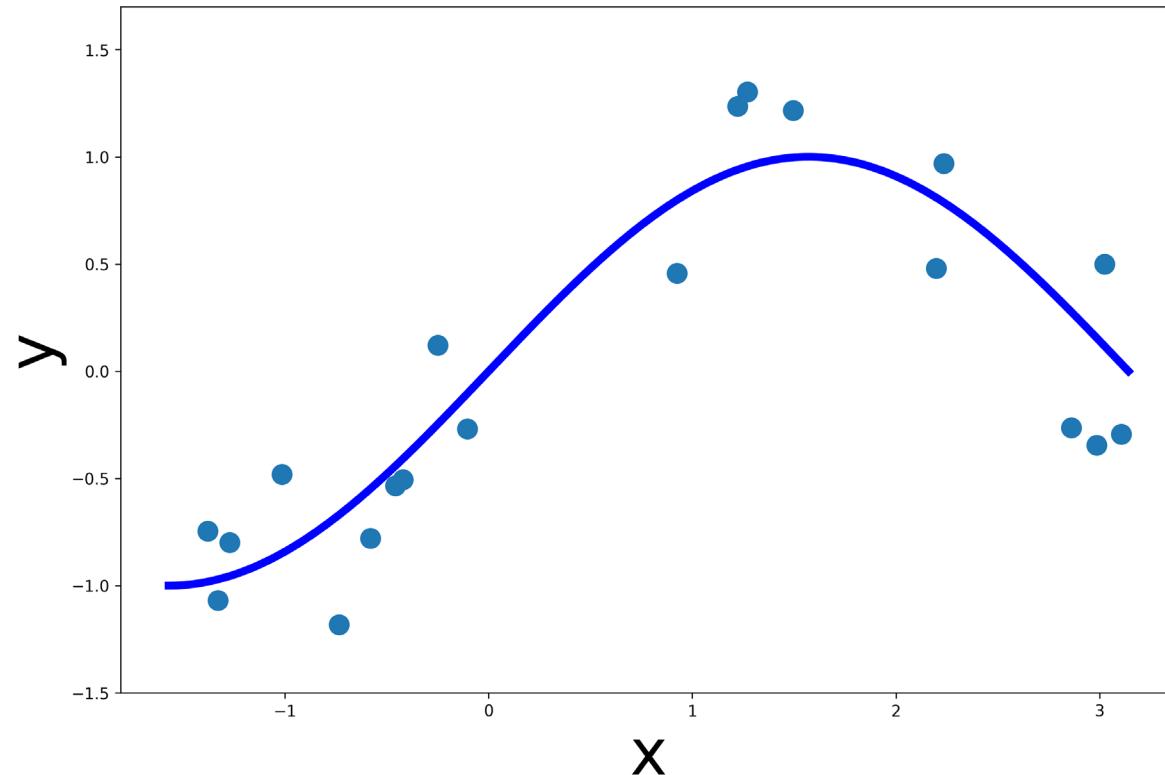
○ «Обучение с учителем»

- восстановление регрессии

**ЧТО Я ХОЧУ?** – значение  $y$

$$y \in \mathbb{R}^m$$

$m$  – размерность целевой переменной



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)

○ «Обучение с учителем»

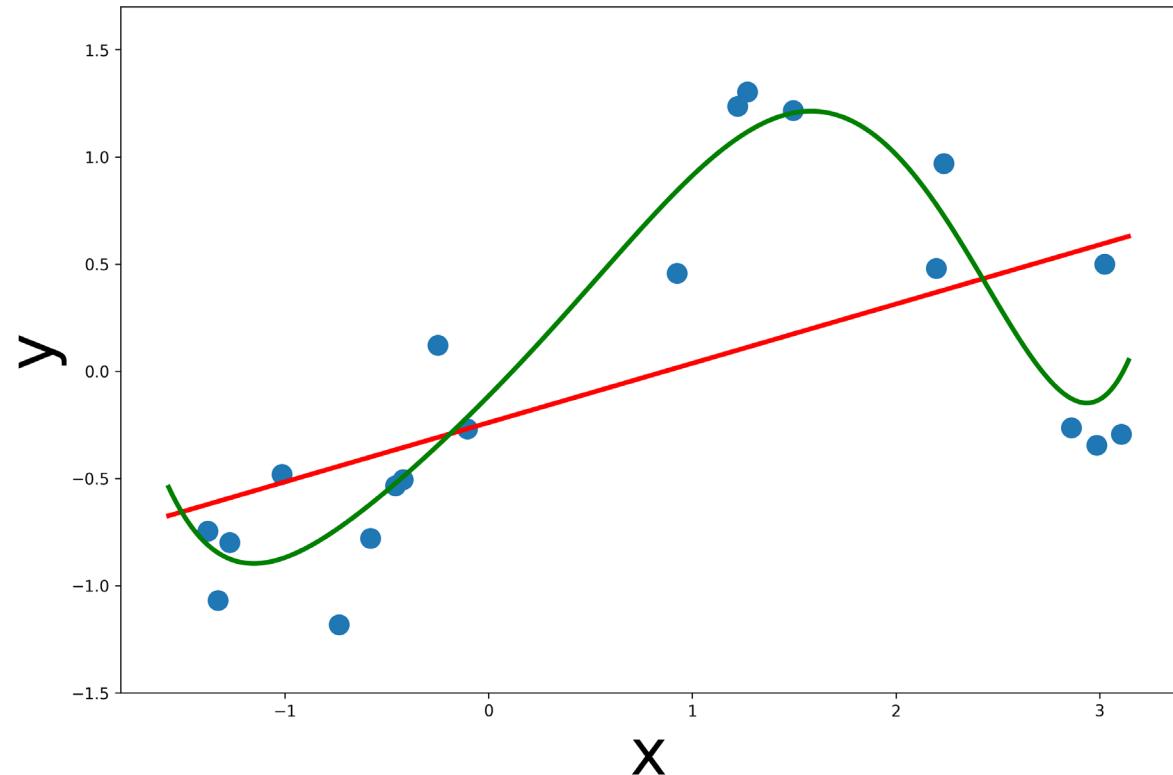
- восстановление регрессии

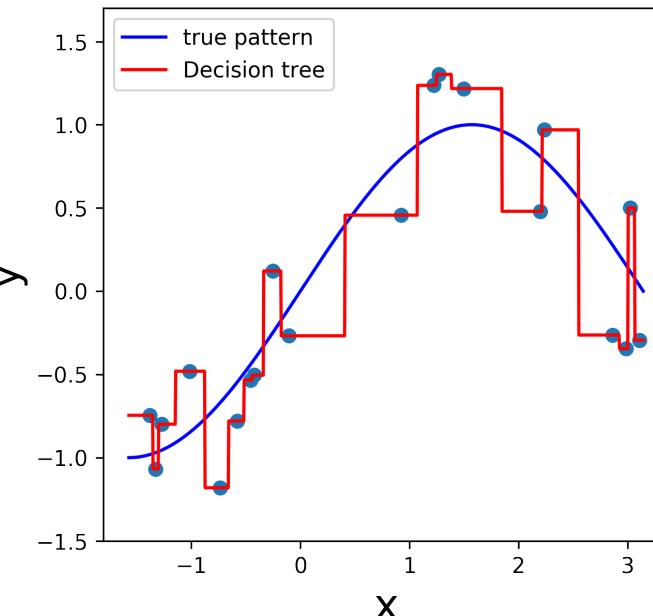
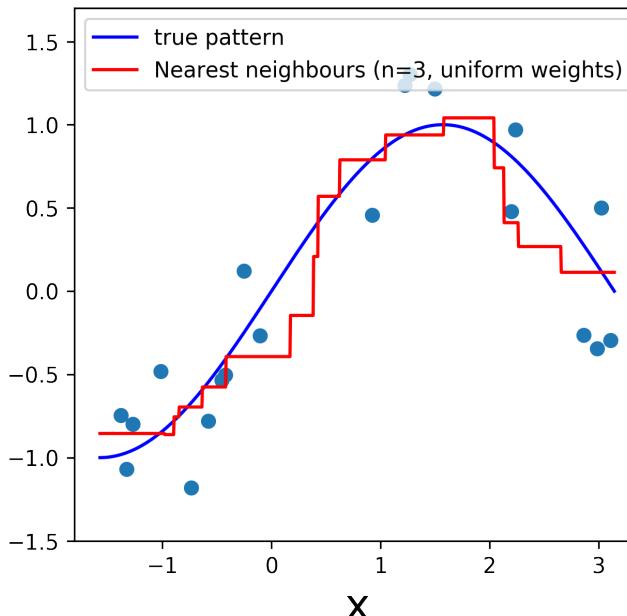
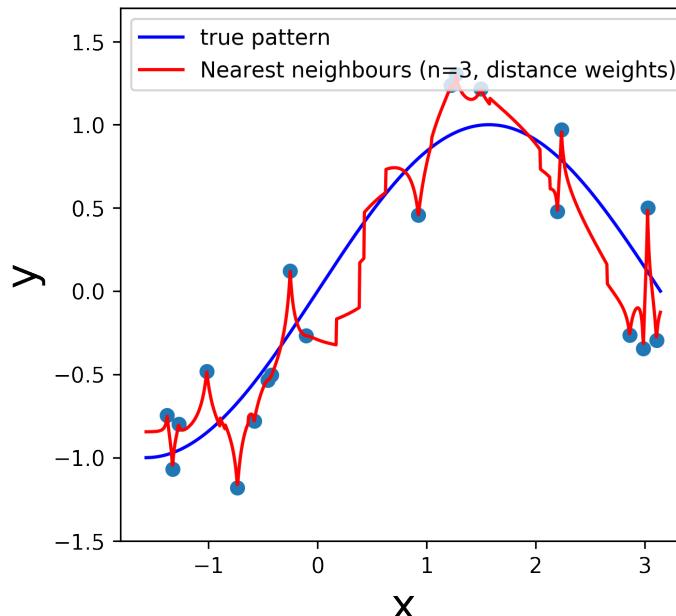
Примеры моделей:

Линейная регрессия

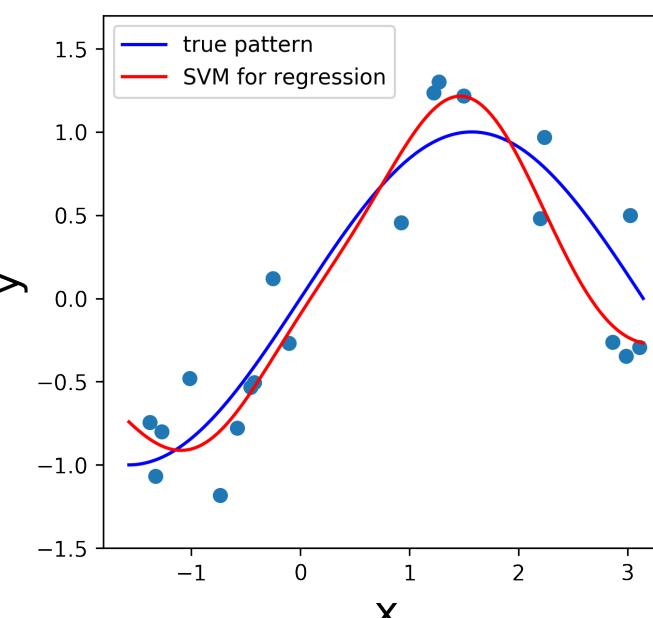
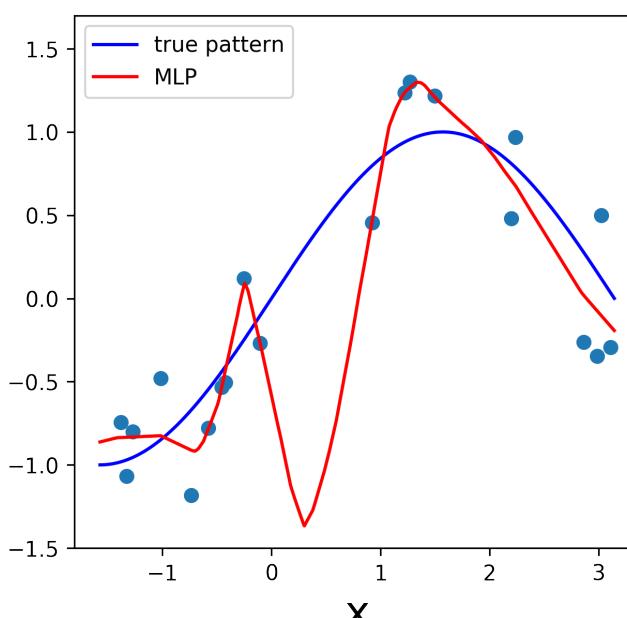
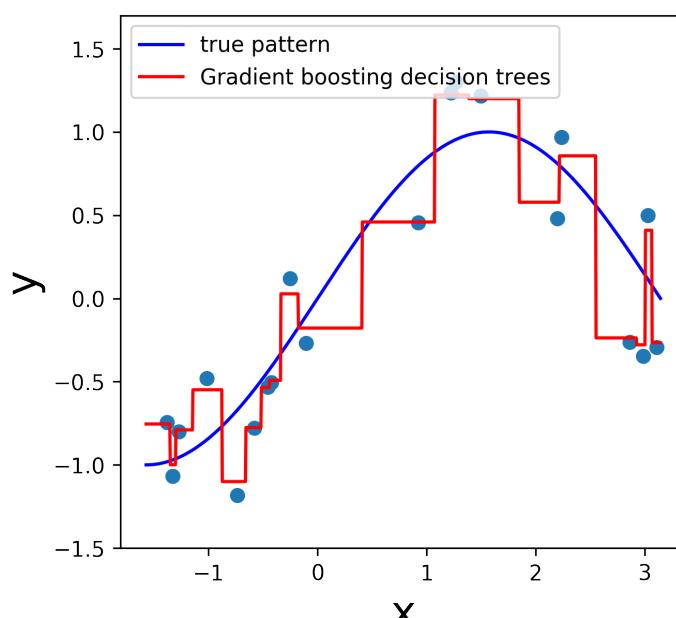
$$\hat{y} = ax + b$$

$$\hat{y} = p^{(6)}(x)$$





1Я)

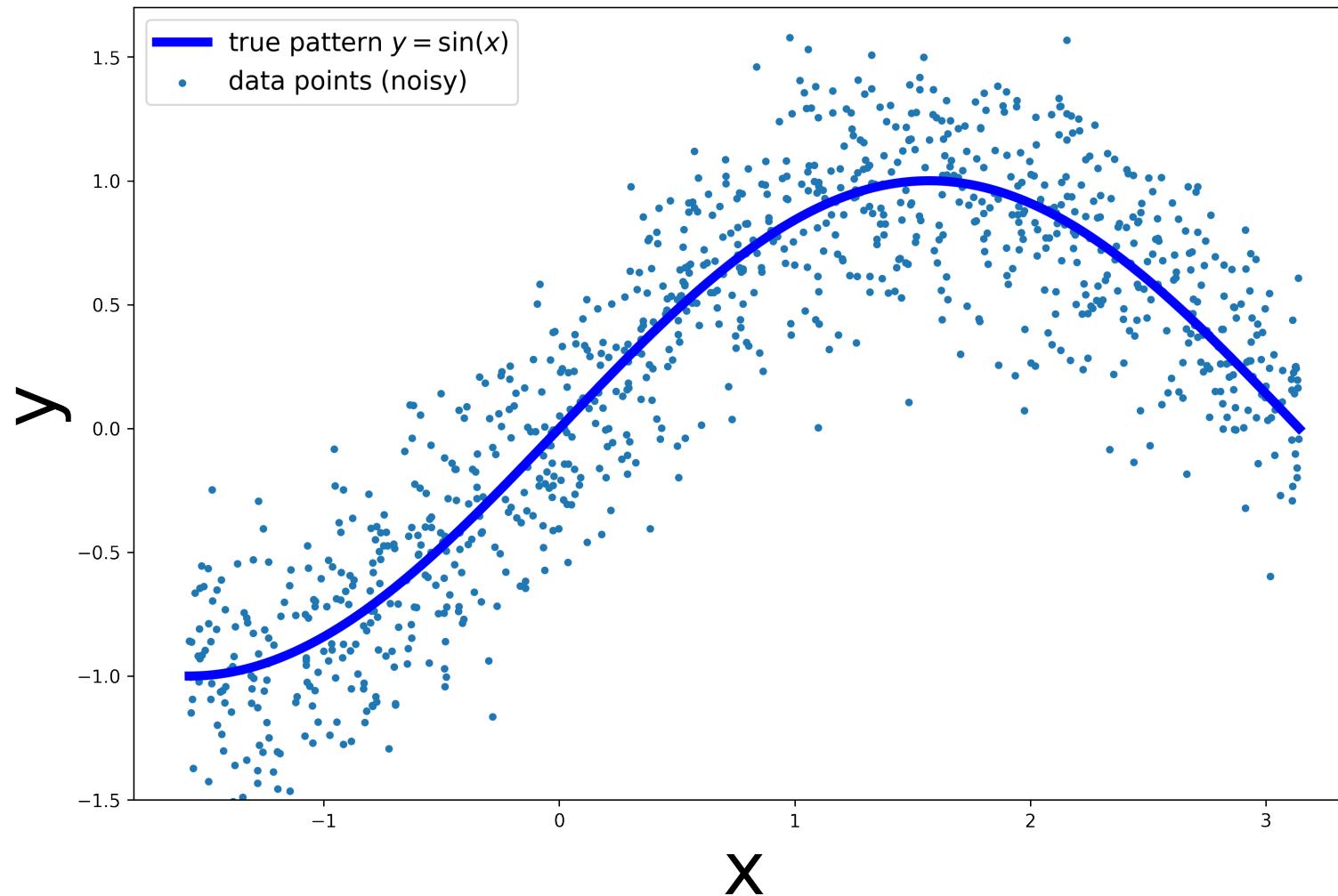


2Я)

3

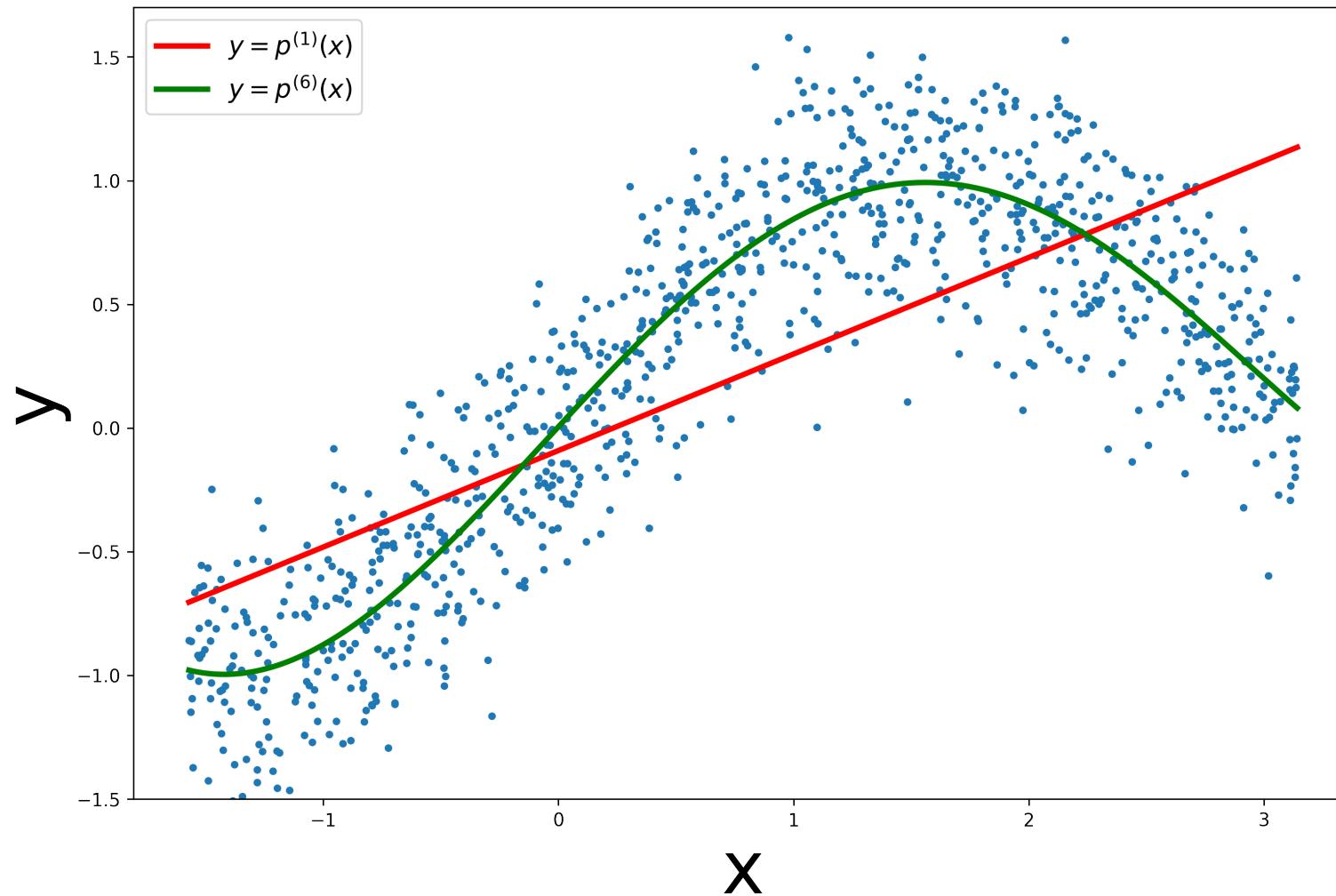
# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

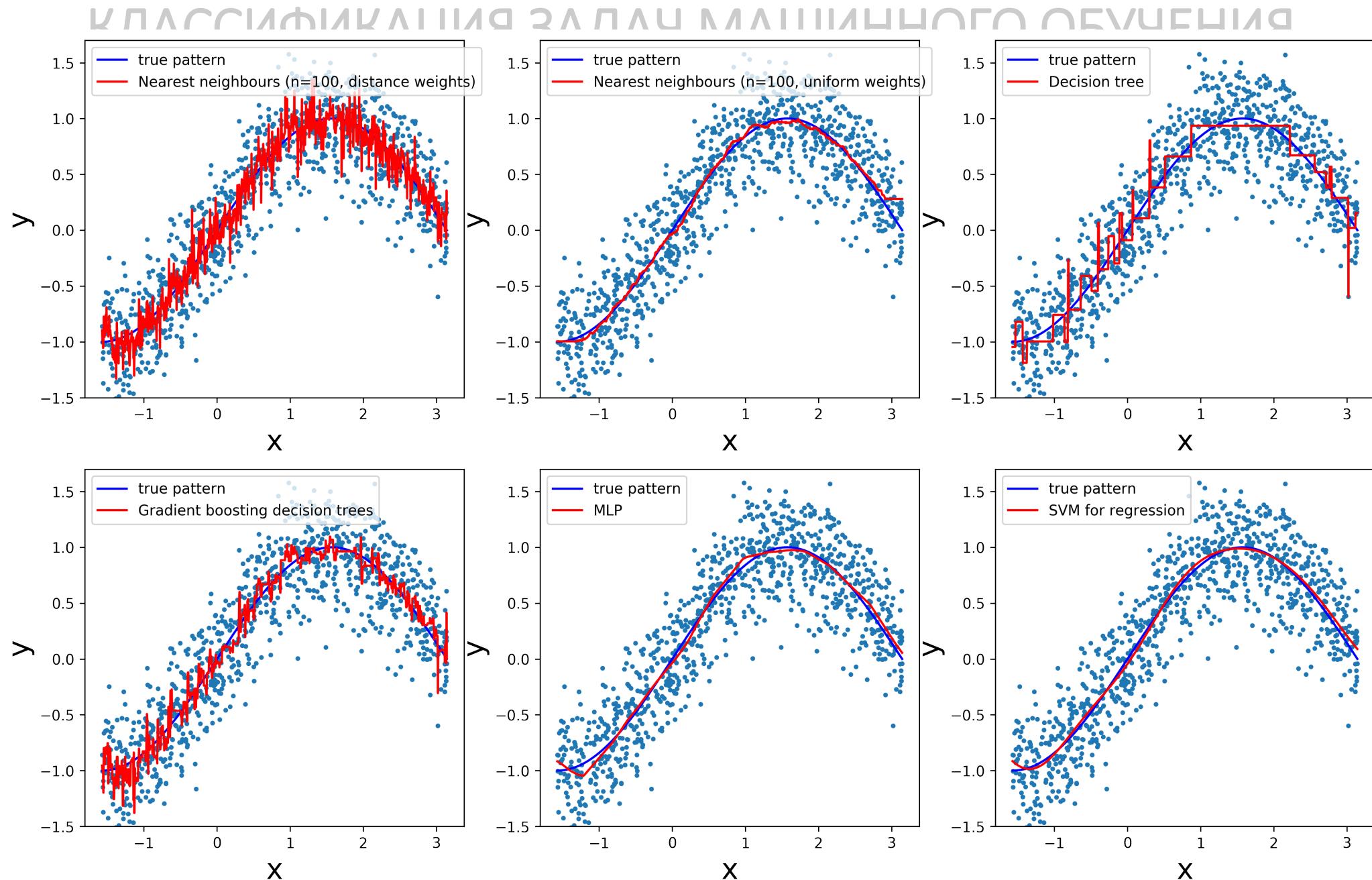
ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)





## ремарки

- Количество размеченных данных (зачастую) играет роль
- Разные модели ведут себя по-разному в зависимости от шума в данных, от количества данных, от наличия выбросов в данных
- Сложная точная модель – не обязательно лучшая для конкретной задачи

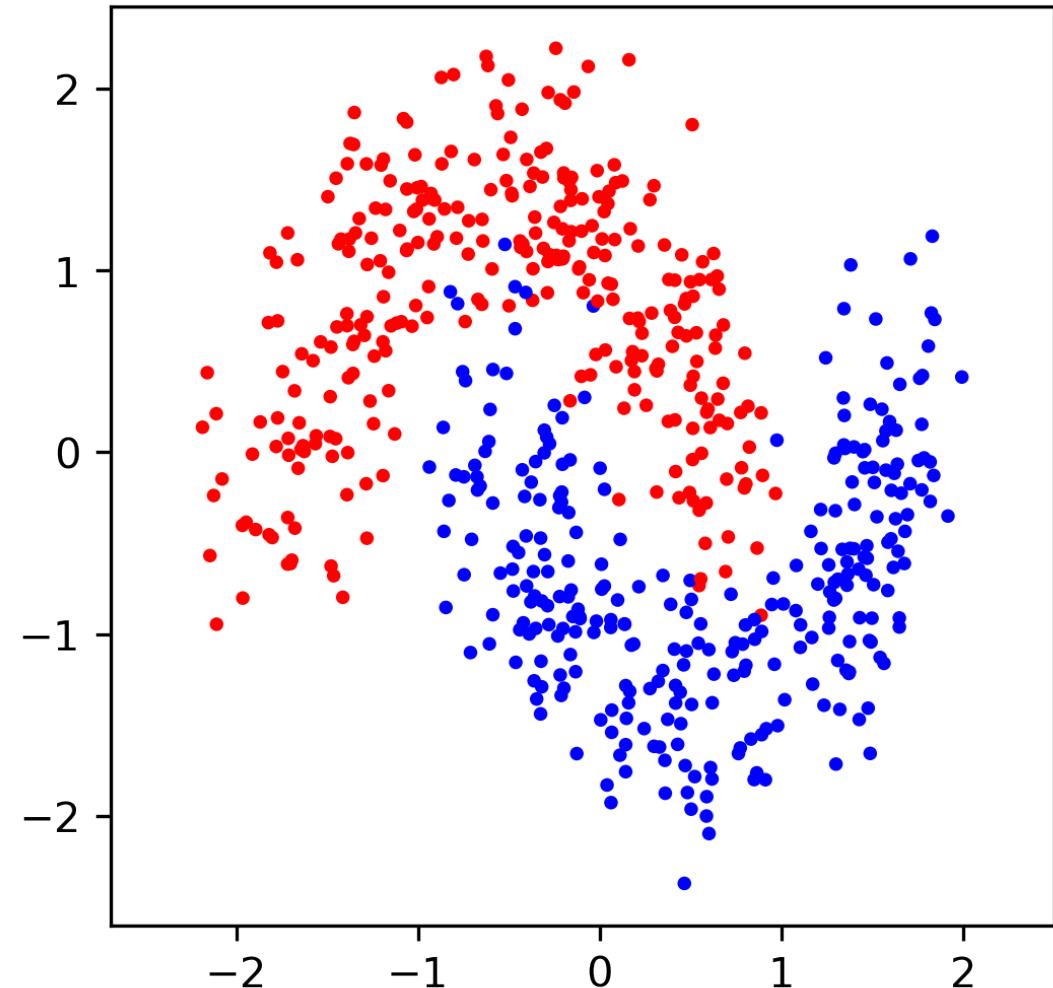
# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: **сформулировать задачу** (в терминах машинного обучения)

- «Обучение с учителем»

- восстановление регрессии
- классификация

**ЧТО Я ХОЧУ?** – метку класса  
**«красный или синий?»**  
(бинарная классификация)



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

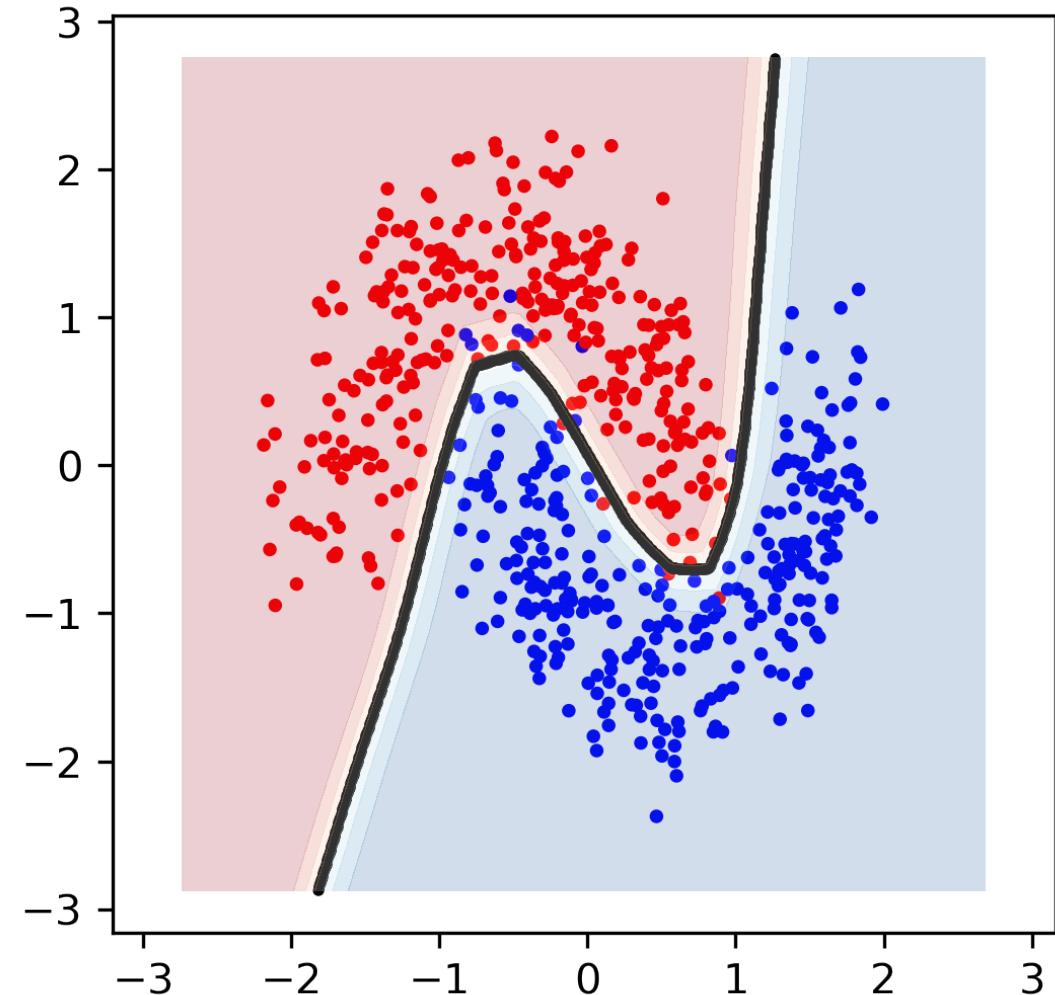
ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)

- «Обучение с учителем»

- восстановление регрессии
- классификация

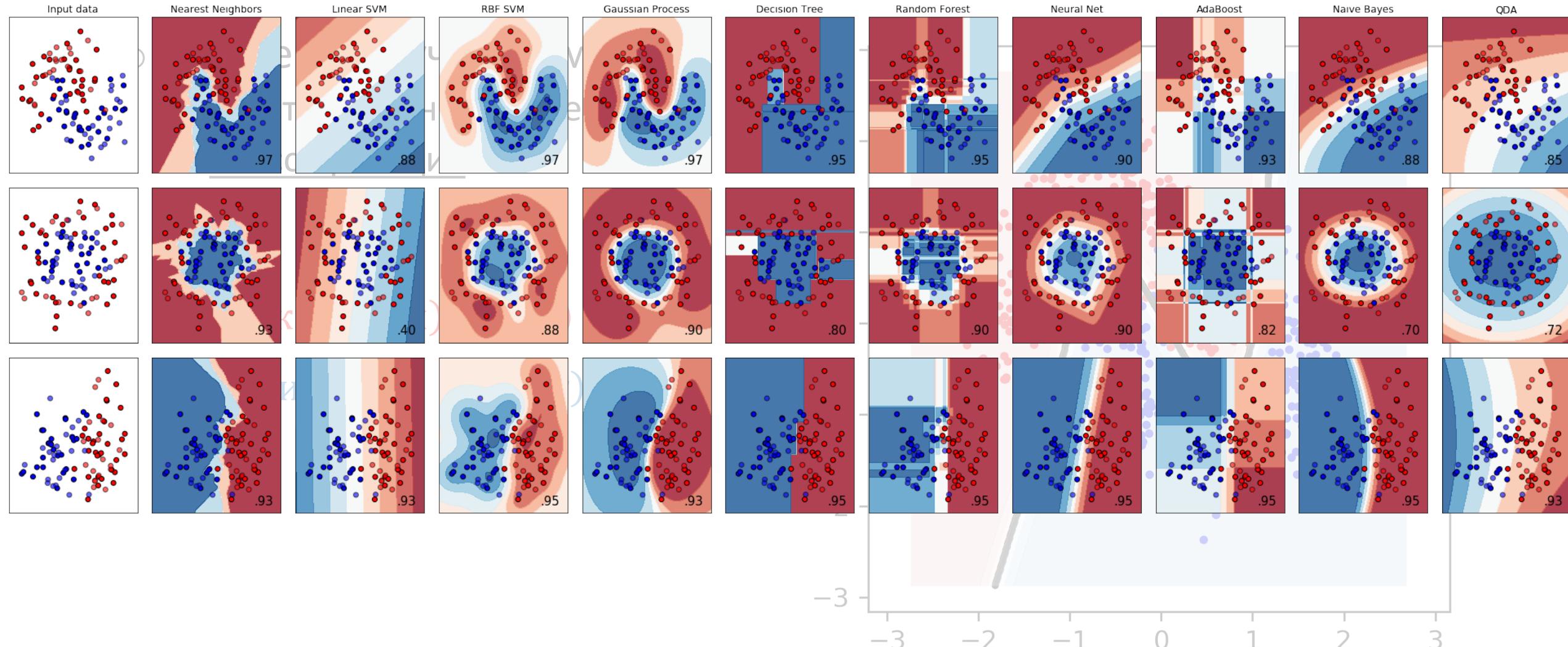
$$\hat{p}(\text{красный}) = f(x)$$

$$\hat{p}(\text{синий}) = 1 - f(x)$$



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

## Сравнение моделей классификации\*



\* [https://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)

# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

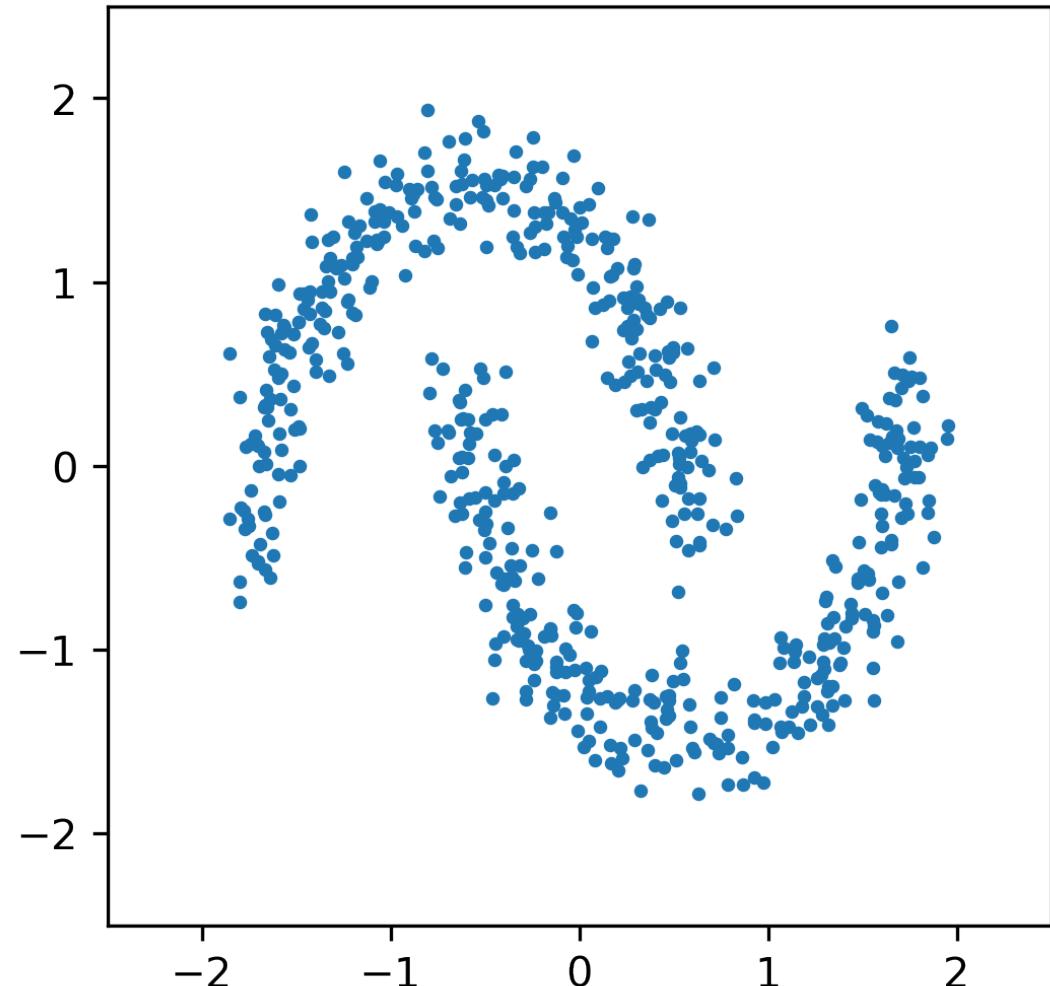
ЦЕЛЬ: **сформулировать задачу** (в терминах машинного обучения)

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - поиск структуры в данных

**что я хочу?**

- метки групп
- знать, есть ли группы?
- сколько групп?



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

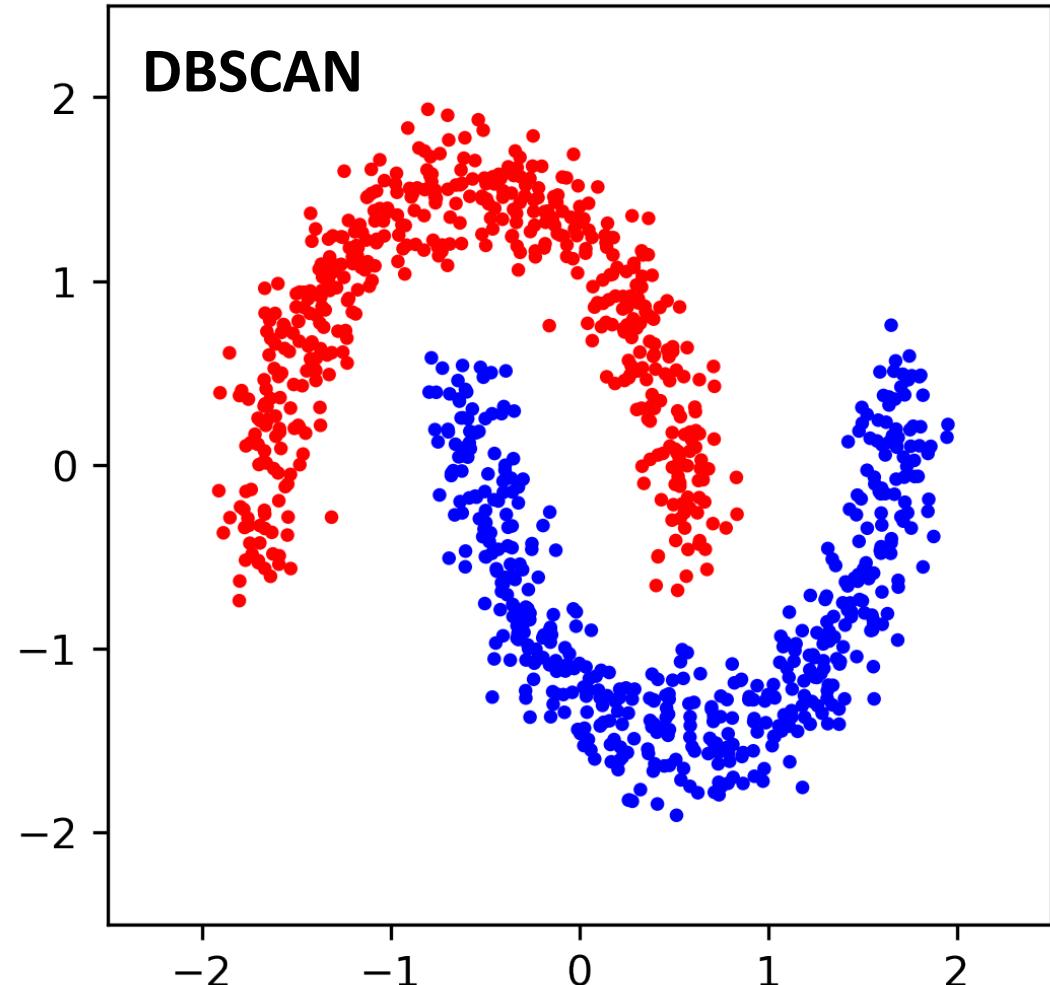
ЦЕЛЬ: **сформулировать задачу** (в терминах машинного обучения)

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

**что я хочу?**

- метки групп
- знать, есть ли группы?
- сколько групп?



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: **сформулировать задачу** (в терминах машинного обучения)

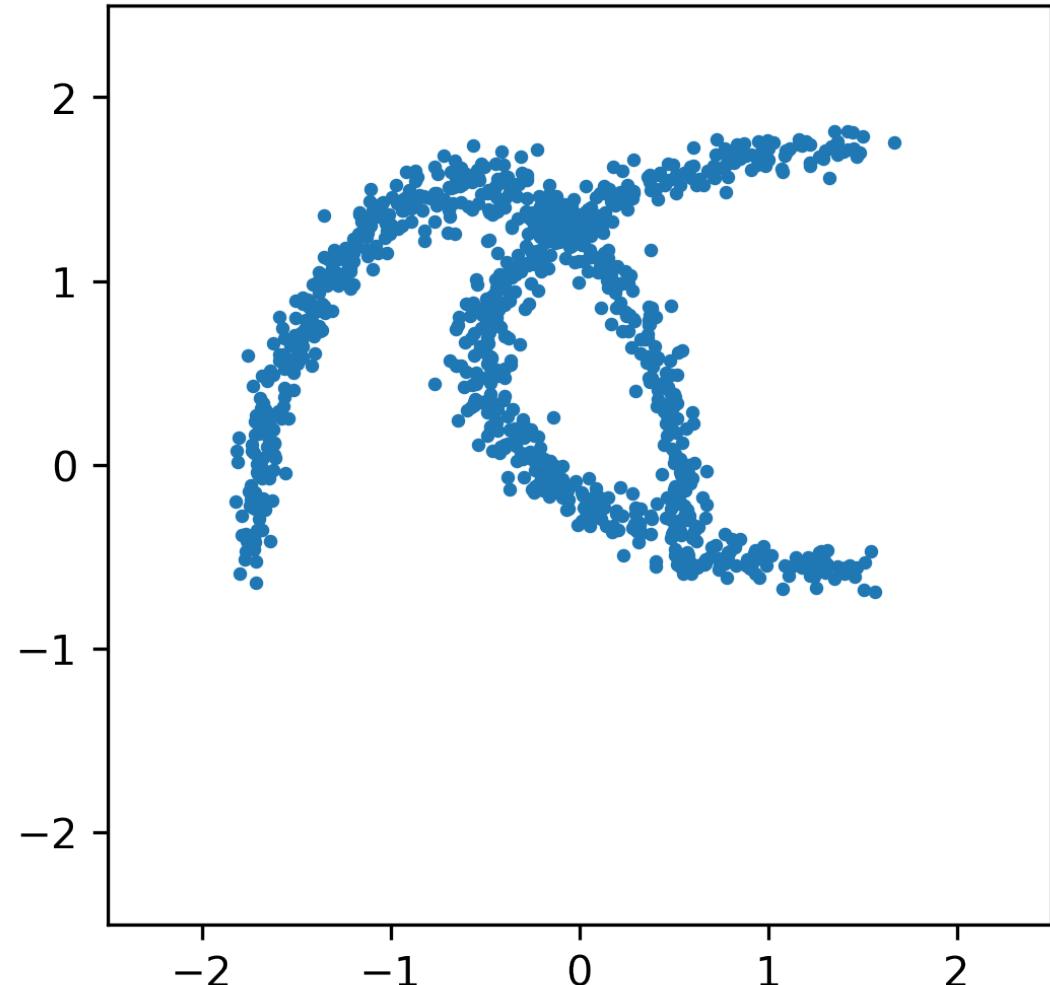
типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

**Всегда ли есть решение?**

хоть какое-нибудь

**ДА**



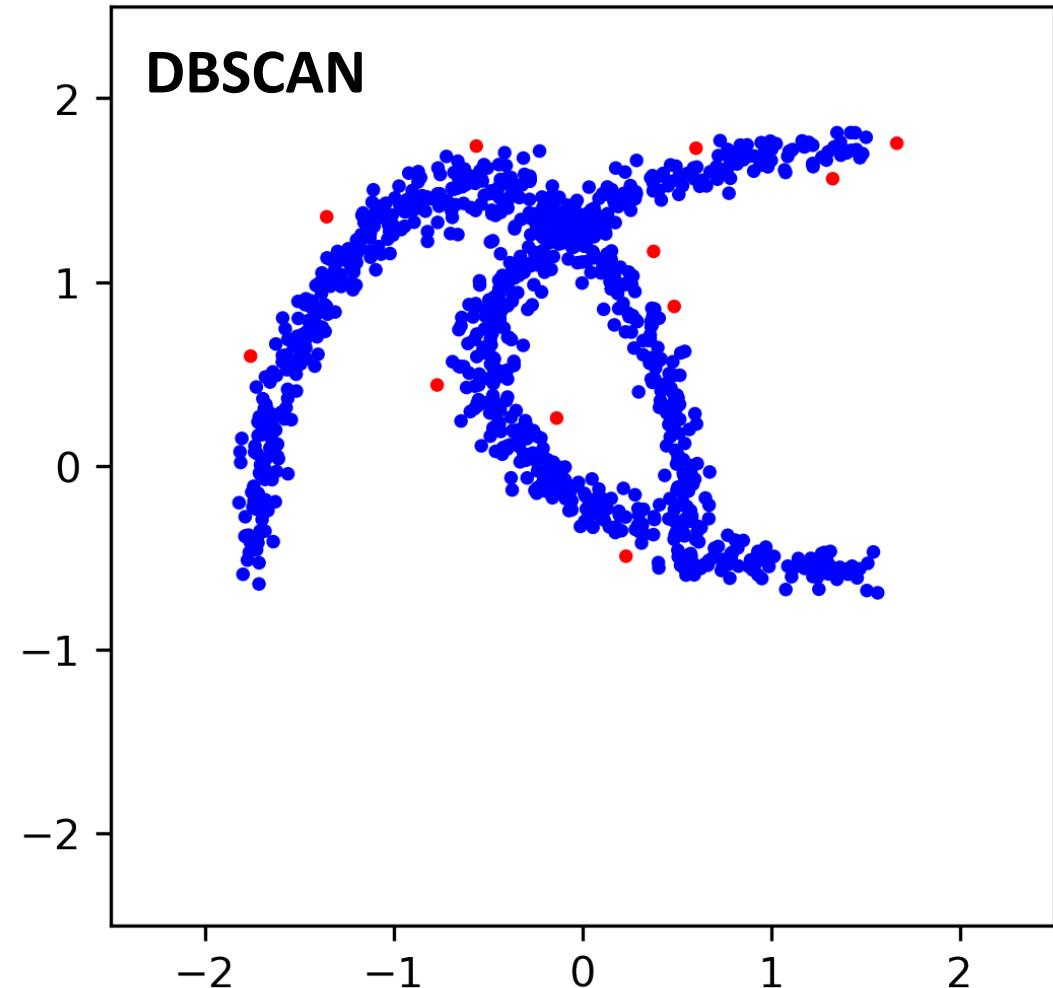
# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: **сформулировать задачу** (в терминах машинного обучения)

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

Всегда ли есть решение,  
**которое мне понравится?**



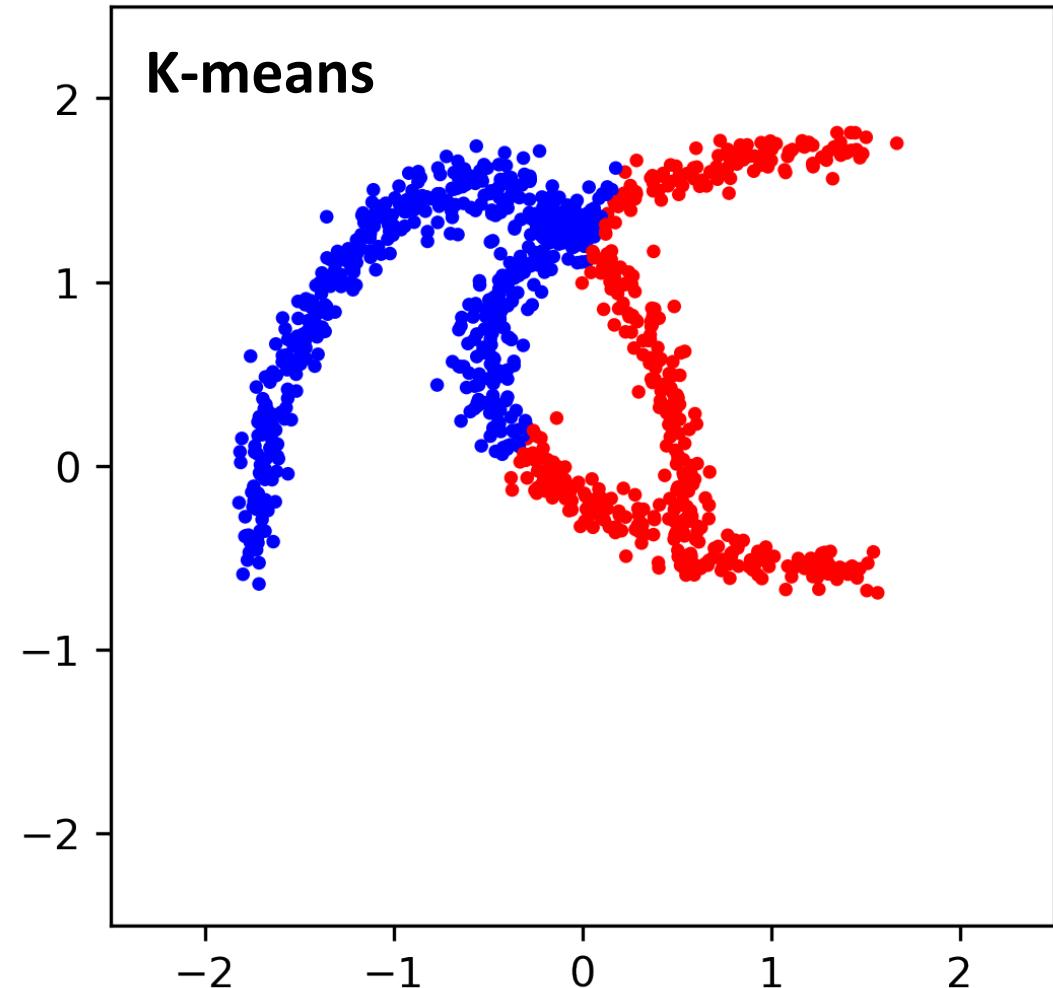
# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: **сформулировать задачу** (в терминах машинного обучения)

типы задач:

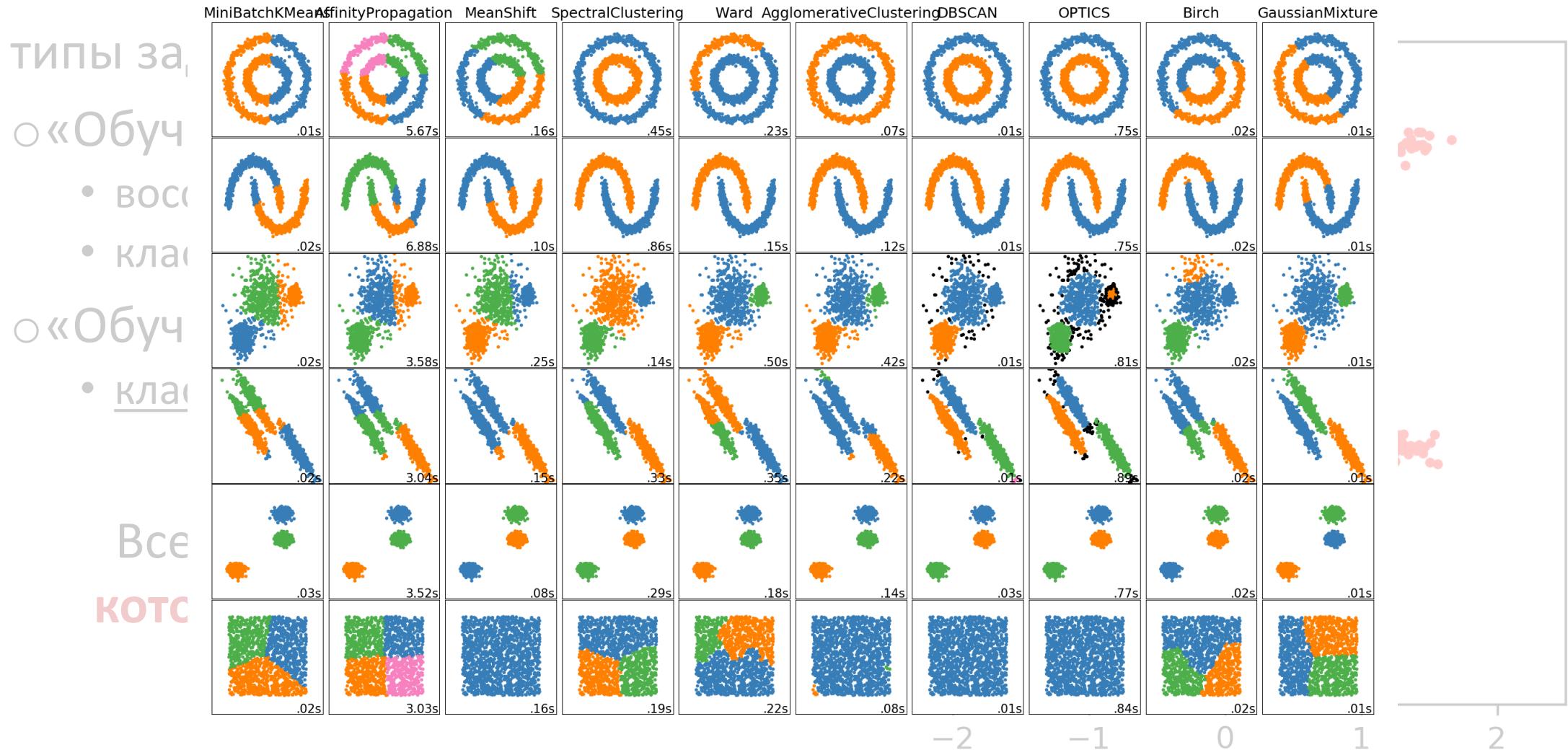
- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

Всегда ли есть решение,  
**которое мне понравится?**



# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

## Сравнение моделей кластеризации\*



## ремарки

Кластеризация:

- Количество данных часто, но не всегда играет роль
- Разные модели ведут себя по-разному в зависимости от шума в данных, от количества данных, от наличия выбросов в данных, от наличия структуры в данных
- Разные модели дают разный результат, но нет «более правильного» результата. Есть «более подходящий» для целей конкретного исследования.

# КЛАССИФИКАЦИЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЦЕЛЬ: сформулировать задачу (в терминах машинного обучения)

типы задач:

○ «Обучение с учителем»

- восстановление регрессии
- классификация

**TO BE CONTINUED**

○ «Обучение без учителя»

- кластеризация
- понижение размерности

что я хочу?

- новое представление (признаковое описание) данных в пространстве меньшей размерности

цели:

- Визуализация на плоскости, в 3D
- Базовое обучением (в контексте т.н. «проклятия размерности»)
- Сжатие данных с минимальными потерями
- Сокращение вычислительных затрат при обработке данных
- Извлечение значимых признаков, feature engineering

Пожелания:

- Сохранение структуры данных
- Сохранение отношений близости между объектами (событиями)
- Возможность визуализации
- Интерпретируемость новых признаков